WHAT WE CLAIM IS:

1. A system for processing audio signals comprising:

(a) a splitter for dividing an input audio signal into a first and one or more secondary signal portions, which in combination provide a complete representation of the input signal, wherein the first signal portion contains information sufficient to reconstruct a representation of the input signal;

(b) a first encoder for providing encoded data about the first signal portion, and one or more secondary encoders for encoding said secondary signal portions, wherein said secondary encoders receive input from the first signal portion and are capable of providing encoded data regarding the first signal portion; and

(c) a data assembler for combining encoded data from said first encoder and said secondary encoders into an output data stream.

2. The system of claim 1 further comprising a decoder for reconstructing the input signal from information in said first signal portion.

3. The system of claim 1 wherein dividing the input signal is done in the frequency domain, and the first signal portion corresponds to the base band of the input signal.

4. The system of claim 1 wherein said signal portions are encoded at sampling rates different from that of the input signal.

5. The system of claim 2 further comprising one or more secondary decoders for decoding information encoded by said secondary encoders.

6. The system of claim 1 wherein said first encoder and said secondary encoders are embedded encoders.

7. The system of claim 1 wherein said splitter is a filter bank.

8. The system of claim 1 wherein said splitter is a Fast Fourier Transform (FFT) computing device.

9. The system of claim 8 wherein said splitter divides the input signal into M octave bands.

10. The system of claim 9 further comprising M1 decoders, $1 \leq M1 \leq M$, for providing an output signal that reconstructs the input signal from information in M1 signal portions of the input signal.

11. The system of claim 10 wherein the output signal has sampling frequency that is 2^{M1} times lower than the sampling frequency of the input signal.

12. The system of claim 1 wherein said output data stream comprises data packets suitable for transmission over a packet-switched network.

13. The system of claim 12 wherein said data packets are prioritized in accordance with the signal portion they represent.

14. The system of claim 12 wherein said data packets are assembled as to represent said two or more signal portions of the input signal.

15. A method for processing audio signals comprising:

(a) dividing an input audio signal into a first and one or more secondary signal portions, which in combination provide a complete representation of the input signal, wherein a first signal portion contains information sufficient to reconstruct a representation of the input signal;

(b) providing first encoded data about the first signal portion, and secondary encoded data about at least one secondary signal portion, wherein said secondary encoded data further comprises information about the first signal portion; and

(c) combining said first encoded data and said secondary encoded data into an output data stream.

16. The method of claim 15 further comprising the step of decoding the output data stream to reconstruct the input signal.

17. The method of claim 15 wherein said signal portions are encoded at sampling rates different from that of the input signal.

18. The method of claim 15 wherein said dividing is performed as a Fast Fourier Transform (FFT) computation.

19. The method of claim 18 further comprising the step of decoding the output data stream using M1 decoders, $1 \leq M1 \leq M$, for providing an output signal that reconstructs the input signal from information in M1 signal portions of the input signal.

20. The method of claim 19 wherein the output signal has sampling frequency that is 2^{M1} times lower than the sampling frequency of the input signal.

21. A system for embedded coding of audio signals comprising:

(a) a frame extractor for dividing an input signal into a plurality of signal frames corresponding to successive time intervals;

(b) means for providing parametric representations of the signal in each frame, said parametric representations being based on a signal model;

(c) means for providing a first encoded data portion corresponding to a user-specified parametric representation, which first encoded data portion contains information sufficient to reconstruct a representation of the input signal;

(d) means for providing one or more secondary encoded data portions of the user-selected parametric representation; and

(e) means for providing an embedded output signal based at least on said first encoded data portion and said one or more secondary encoded data portions of the user-selected parametric representation.

22. The system of claim 21 further comprising:

(f) means for providing representations of the signal in each frame, which are not based on a signal model.

23. The system of claim 22 further comprising

(g) means for selecting a specific one from the representations in (b) and (f) based on user-selected constraints.

24. The system of claim 21 wherein said means for providing parametric representations of the signal in each frame comprises a pitch detector for computing a first estimate of the pitch of a signal in each frame; means for determining parameters of sinusoids representing the signal in each frame; and a spectrum envelope encoder for encoding the shape of the envelope of the signal in each frame.

25. The system of claim 21 wherein said means for providing an embedded output signal comprises a bit stream assembler for providing an output bit stream containing user-specified information about parameters of at least one sinusoid in the spectrum of the input signal, and about parameters representing a spectrum envelope of the signal in each frame.

26. The system of claim 21 further comprising means for decoding the embedded output signal.

27. The system of claim 26 wherein said means for decoding operate at a sampling frequency different from a sampling frequency of the input signal.

28. The system of claim 21 wherein said means for providing an embedded output signal comprises means for assembling data packets suitable for transmission over a packet-switched network.

29. A method for multistage vector quantization of signals comprising:

(a) passing an input signal through a first stage of a multistage vector quantizer having a predetermined set of codebook vectors, each vector corresponding to a Voronoi cell, to obtain error vectors corresponding to differences between a codebook vector and an input signal vector falling within a Voronoi cell;

(b) determining probability density functions (pdfs) for the error vectors in at least two Voronoi cells;

(c) transforming error vectors using a transformation based on the pdfs determined for said at least two Voronoi cells; and

(d) passing transformed error vectors through at least a second stage of the multistage vector quantizer to provide a quantized output signal.

30. The method of claim 29 further comprising the step of performing an inverse transformation on the quantized output signal to reconstruct a representation of the input signal.

31. The method of claim 29 wherein in step (c) the transformation comprises scaling the sizes of said at least two Voronoi cells as to approximately equalize these sizes.

32. The method of claim 31 wherein scaling factor for a Voronoi cell is determined as the inverse of an average for the Euclidean distance between the codebook vector for the Voronoi cell and a set of training vectors.

33. The method of claim 29 wherein in step (c) the transformation comprises rotating the error vector at an angle, which is determined by the Voronoi cell.

34. The method of claim 33 wherein the rotation angle is determined as the angle between the codebook vector for the Voronoi cell and one of the coordinate axes of the cell.

35. The method of claim 29 wherein in step (c) the transformation comprises both scaling and rotating the error vector at given angle.

36. The method of claim 29 wherein in step (c) a transformation for inner Voronoi cells is different a transformation for outer Voronoi cells.

37. The method of claim 29 wherein in step (c) the transformation is performed using tuning of translation and rotation parameters as to maximally align boundaries of scaled Voronoi regions and slopes of pdfs in each Voronoi region.

38. A system for processing audio signals comprising;

(a) a frame extractor for dividing an input audio signal into a plurality of signal frames corresponding to successive time intervals;

(b) a frame mode classifier for determining if the signal in a frame is in a transition state;

(c) a processor for extracting parameters of the signal in a frame receiving input from said classifier, wherein for frames the signal of which is determined to be in said transition state said extracted parameters include phase information; and

(d) a multi-mode coder in which extracted parameters of the signal in a frame are processed in at least two distinct paths dependent on whether the frame signal is determined to be in a transition state.

39. The system of claim 38 wherein said extracted parameters comprise gain, pitch and voicing parameters and parameters related to Linear Prediction Coefficients (LPCs).

$$y(n;\omega_0) = \mu\sum_{k=1}^{K}\gamma_k\exp(jn\omega_0)+\sum_{l=1}^{L}\sum_{k=1}^{K-l}\gamma_{k+l}\gamma_k^*\exp(jnl\omega_0)$$

40. The system of claim 38 wherein said frame mode classifier receives input from said processor for extracting parameters and outputs at least one state flag.

41. The system of claim 40 wherein the multi-mode coder determines one of said at least two distinct processing paths on the basis of said at least one state flag.

42. The system of claim 38 further comprising a decoder for decoding signals in at least two distinct processing paths.

43. The system of claim 38 wherein said distinct processing paths include distinct bit allocation for frames determined to be in different states.

44. A system for processing audio signals comprising:

(a) a frame extractor for dividing an input signal into a plurality of signal frames corresponding to successive time intervals;

(b) means for providing a parametric representation of the signal in each frame, said parametric representation being based on a signal model;

(c) a non-linear processor for providing refined estimates of parameters of the parametric representation of the signal in each frame; and

(d) means for encoding said refined parameter estimates.

45. The system of claim 44 wherein said refined estimates comprises an estimate of the pitch.

46. The system of claim 44 wherein said refined estimates comprises an estimate of a voicing parameter for the input speech signal.

47. The system of claim 44 wherein said refined estimates comprises an estimate of a pitch onset time for an input speech signal.

48. The system of claim 44 wherein said non-linear processor computes the maximum of a correlation function of the input signal over a set of complex frequencies.

49. The system of claim 45 wherein the computation is done iteratively.

50. The system of claim 44 wherein a measure of voicing for the input signal is computed as

$$\rho(\omega_0) = \sum_{m=1}^{M} |Y_m|^2 \; 0.5*[1+\cos(2\pi\omega_m/\omega_0)] / \sum_{m=1}^{M} |Y_m|^2$$

where Y_m are complex amplitudes of the output of a nonlinear operation defined over the input signal s(n) as defined

$$\begin{aligned} y(n) &= \mu \sum_{k=1}^{K} s_k(n) + \sum_{l=1}^{L} \sum_{k=1}^{K-l} s_{k+l}(n)\, s_k^*(n) \\ &= \mu \sum_{k=1}^{K} \gamma_k \exp(jn\omega_k) + \sum_{l=1}^{L} \sum_{k=1}^{K-l} \gamma_{k+l}\gamma_k^* \exp[jn(\omega_{k+l}-\omega_k)] \end{aligned} \tag{1}$$

where $\gamma_k = A_k \exp(j\theta_k)$ is the complex amplitude and where $0 \leq \mu \leq 1$ is a bias factor.